Problem 1. Contestant should change to door B.

Let $A, B, C$ be the events prize is behind door $A, B, C$ (respectively).
Let $\text{Open } B, \text{Open } C$ be the events host opens door $B, C$ (respectively).

Before host opens door

$P(A) = P(B) = P(C) = \frac{1}{3}$ the prior.

$p(\text{Open } B | A) = \frac{1}{2}$, $P(\text{Open } C | A) = \frac{1}{2}$ likelihood

$p(\text{Open } B | B) = 0$, $P(\text{Open } C | B) = 1$ observation.

$p(\text{Open } B | C) = 1$, $P(\text{Open } C | C) = 0$.

\[ p(A | \text{open } C) = \frac{P(\text{open } C | A) P(A)}{P(\text{open } C)} = \frac{1}{6} \cdot \frac{1}{3} = \frac{1}{18} \]

\[ p(B | \text{open } C) = \frac{P(\text{open } C | B) P(B)}{P(\text{open } C)} = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9} \]

\[ p(C | \text{open } C) = \frac{P(\text{open } C | C) P(C)}{P(\text{open } C)} = 0 \]

Normalization gives:

$p(A | \text{open } C) = \frac{1}{3}$ So contestant

$p(B | \text{open } C) = \frac{2}{3}$ should switch to door B.

Note: Contestant guesses door A, Host opens door C.
Problem 2. Image I, \( \omega \in \{S, P\} \)

Decision Rule \( \hat{\omega}(I) \in \{S, P\} \)

Loss Function \( L(\hat{\omega}(I), \omega) \)

\( \omega \), \( I \)

Risk \( R(\hat{\omega}) = \sum_{\omega} L(\hat{\omega}(I), \omega) \cdot P(I|\omega) \)

Bayes Rule \( \hat{\omega} = \arg\min_{\omega} R(\omega) \).

Equivalently, \( \hat{\omega}(I) = \arg\min_{\omega} \sum_{\omega} L(\omega(I), \omega) \cdot P(I|\omega) \)

Bayes Risk \( \hat{R} = \min_{\omega} R(\omega) = R(\hat{\omega}) \) (almost always).

If errors are weighted equally, then
\[
L(\omega(I), \omega) = 1 - \delta_{\omega(I), \omega} \Rightarrow \begin{cases} 0, & \text{if } \omega(I) = \omega \\ 1, & \text{if } \omega(I) \neq \omega \end{cases}
\]

Kronecker Delta.

Set of samples \( \{ (I_\mu, \omega_\mu) : \mu = 1 \to N \} \)

Assume these are i.i.d. samples from some unknown distribution \( P(I|\omega) \).

Empirical Risk \( R_{emp}(\hat{\omega}) = \frac{1}{N} \sum_{\mu=1}^{N} L(\hat{\omega}(I_\mu), \omega_\mu) \)

As \( N \to \infty \), \( R_{emp}(\omega) \to R(\omega) \)
Problem 2. (cont.)

Two Strategies:

(A) Extreme Generative: Learn the probability distribution \( P(I_i, w) = P(I_i|w) P(w) \). Then seek decision rule \( \hat{I} \) to minimize Bayes risk.

(B) Extreme Discriminative: Start from the empirical risk, seek a decision rule to minimize \( R_{emp} \). Often restrict the form of the decision rule for computational efficiency and to prevent overgeneralizing.

Generalization: require the decision rule to still be effective on testing data - i.e. data which isn't in the training data, but which is sampled from the same distribution \( P(I_i|w) \).

Memorization - the "intelligent paint" strategy
Good performance on training data, but poor performance on testing data.

Moral: the complexity of the model depends on the amount of data.
Complex Model + Little Data = Bad Generalization.
Problem 3. \[
P(x,y \mid w) = \frac{1}{2\pi \sigma^2} e^{-\frac{(x-\mu_x, y-\mu_y)^2}{2\sigma^2}}
\]
(20 pts)
Prior: \[p(w_1) = 4p(w_2)\]

hence \[p(w_1) = \frac{4}{5}\]
\[p(w_2) = \frac{1}{5}\]

Loss function weights all errors equally. Hence Bayes Rule selects \(w_1\) to maximize \(p(w \mid xy) = \frac{p(xy \mid w)p(w)}{p(xy)}\).

Decision Boundary is at \(p(w_1 \mid xy) = p(w_2 \mid xy)\)
\[
\log p(xy \mid w_1) + \log p(w_1) = \log p(xy \mid w_2) + \log p(w_2)
\]
\[
\frac{(x-\mu_{x1})^2 + (y-\mu_{y1})^2}{2\sigma_{1}^2} + \log \sigma_{1}^2 - \log 4 = \frac{(x-\mu_{x2})^2 + (y-\mu_{y2})^2}{2\sigma_{2}^2} + \log \sigma_{2}^2
\]

Case 1: \(\sigma_{1}^2 = \sigma_{2}^2 = 1\).

After some algebra, this simplifies to a linear boundary \(y = x - \frac{1}{2} \log 2\).

Points above boundary \(y > x - \frac{1}{2} \log 2\) are classified as \(w_1\), points below are classified as \(w_2\) (Note: boundary is closer to \(w_2\) because the prior biases towards \(w_2\)).

Case 2: \(\sigma_{1}^2 = 4, \sigma_{2}^2 = 16\). After algebra, this simplifies to a circular boundary:
\[
3x^2 + 3(y-10)^2 = 96 + 16y + 128\log 2
\]
Problem 4.
Two Class Discrimination

Decision Rule \( \omega_1 \) if \( \log \frac{p(x|\omega_1)}{p(x|\omega_2)} > T \)
\( \omega_2 \) otherwise.

True +ve's \( \sum p(x|\omega_1) \)
False +ve's \( \sum p(x|\omega_2) \)
\( x : \log \frac{p(x|\omega_1)}{p(x|\omega_2)} > T \)
\( x : \log \frac{p(x|\omega_1)}{p(x|\omega_2)} < T \)

ROC curve: plot the true positives \( \sum p(x|\omega_1) \)
against the false +ve's as the threshold \( T \) varies.

For Gaussians with identical variances, the decision boundary is at:
\( \log \frac{p(x|\mu_1,\sigma^2)}{p(x|\mu_2,\sigma^2)} = T \)
\( = \frac{(x-\mu_1)^2 + (x-\mu_2)^2}{2 \sigma^2} - T = 0 \)
\( x_T = \frac{\mu_1 + \mu_2}{2} + \frac{T \sigma^2}{2} \)

ROC curve is of form
At \( T = \infty \) all points are classified as \( \omega_2 \)
At \( T = \infty \) all points are classified as \( \omega_1 \) so
True +ve = False +ve's = 0
Problem 5.

Entropy of \( p_i(x) \):

\[
\text{Ent}_p(x) = -\int_{-\infty}^{\infty} p_i(x) \log p_i(x) \, dx \quad \text{(Integral for continuous variable)}
\]

\[
= \int_{-\infty}^{\infty} p_i(x) \left\{ \frac{(x-\mu_i)^2 + \log \sqrt{2\pi\sigma_i^2}}{2\sigma_i^2} \right\} \, dx
\]

\[
= \frac{1}{2} \log \sqrt{2\pi\sigma_i^2} \quad \text{(Because} \int_{-\infty}^{\infty} p_i(x)(x-\mu_i)^2 \, dx = \sigma_i^2)\]

Difference in Entropy of \( p_1 \) and \( p_2 \) is \( 0 \),

(because both have the same variance).

Kullback-Leibler: \( D(p_1||p_2) = \int_{-\infty}^{\infty} p_1(x) \log \frac{p_1(x)}{p_2(x)} \, dx \)

\[
= \int_{-\infty}^{\infty} p_1(x) \log p_1(x) \, dx - \int_{-\infty}^{\infty} p_1(x) \log p_2(x) \, dx
\]

\[
= \{-\frac{1}{2} \log \sqrt{2\pi\sigma_1^2} + \int_{-\infty}^{\infty} p_1(x) \left\{ \frac{(x-\mu_1)^2 + \log \sqrt{2\pi\sigma_1^2}}{2\sigma_1^2} \right\} \, dx \}
\]

\[
= -\frac{1}{2} \log \sqrt{2\pi\sigma_1^2} + \int_{-\infty}^{\infty} p_1(x) \left\{ \frac{(x-\mu_1)^2}{2\sigma_1^2} - \frac{2\mu_1x}{2\sigma_1^2} + \frac{\mu_1^2}{2\sigma_1^2} \right\} \, dx
\]

\[
= -\frac{1}{2} \log \sqrt{2\pi\sigma_1^2} + \int_{-\infty}^{\infty} p_1(x) \left\{ \frac{(x-\mu_2)^2}{2\sigma_2^2} - \frac{2\mu_2x}{2\sigma_2^2} + \frac{\mu_2^2}{2\sigma_2^2} \right\} \, dx
\]

\[
= -\frac{1}{2} \log \sqrt{2\pi\sigma_1^2} + \frac{(\mu_2^2 + \mu_1^2)}{2\sigma_2^2} - \frac{2\mu_1\mu_2}{2\sigma_2^2} + \frac{\mu_2^2}{2\sigma_2^2} + \frac{\mu_1^2}{2\sigma_2^2}
\]

\[
= \frac{(\mu_1 - \mu_2)^2}{2\sigma_2^2}
\]

Note: we have used \( \int_{-\infty}^{\infty} p_1(x) x \, dx = \mu_1^2 + \sigma_1^2 \)

\[
\int_{-\infty}^{\infty} p_1(x) \, dx = \mu_1
\]

\[
\int_{-\infty}^{\infty} p_1(x) \, dx = 0
\]
Problem 6. Maximum Entropy Principle:

\[- \int_0^\infty p(x) \log p(x) \, dx + \lambda \left( \int_0^\infty p(x) \, dx - 1 \right) - \lambda \left( \int_0^\infty x p(x) \, dx - \mu \right) \]

Lagrange multiplier

Taking

\[ \frac{S}{\partial \log p(x)} = -1 - \log p(x) + \lambda - \lambda x \]

\[ \Rightarrow \hat{p}(x) = e^{-\lambda x} e^{-1} \]

Taking

\[ \frac{d}{d\lambda} \Rightarrow \int_0^\infty \hat{p}(x) \, dx = 1 \]

\[ \int_0^\infty e^{-\lambda x} \, dx = \left[ -\frac{1}{\lambda} e^{-\lambda x} \right]_0^\infty = \frac{1}{\lambda} \]

hence \[ \hat{p}(x) = \frac{1}{\lambda} e^{-\lambda x} \Rightarrow \text{normalized}. \]

Taking

\[ \frac{d}{d\mu} \Rightarrow \int_0^\infty x \hat{p}(x) \, dx = \mu \]

\[ \int_0^\infty x e^{-\lambda x} \, dx = \left[ -x e^{-\lambda x} \right]_0^\infty + \int_0^\infty e^{-\lambda x} \, dx = \frac{\mu}{\lambda} \]

hence \[ \frac{\mu}{\lambda} = \mu \Rightarrow \lambda = \frac{\mu}{\lambda}. \]

\[ \hat{p}(x) = \frac{1}{\mu} e^{-\lambda x} \mu \]

\[ P(x; \text{HOMY}) = \prod_{i=1}^m e^{-\lambda x_i} \frac{\lambda}{\mu} x_i \]

Sufficient statistic is \[ \lambda \sum_{i=1}^m x_i = \mu \text{ (or x)} \]

To estimate \( \lambda \) by ML, set \( \frac{\partial}{\partial \lambda} \left( \mu \sum_{i=1}^m e^{-\lambda x_i} \right) = 0 \Rightarrow \lambda = \frac{1}{\sum_{i=1}^m x_i} \)